

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Theoretical Computer Science 356 (2006) 245–254

Theoretical  
Computer Science[www.elsevier.com/locate/tcs](http://www.elsevier.com/locate/tcs)

# Completing prefix codes in submonoids

Jean Néraud

Université de Rouen, Faculté des Sciences, Campus du Madrillet, lifar/Département Informatique, Avenue de l'Université,  
BP 12, 76801 Saint Etienne du Rouvray cedex, France

## Abstract

Let  $M$  be a submonoid of the free monoid  $A^*$ , and let  $X \subseteq M$  be a variable length code (for short a code).  $X$  is weakly  $M$ -complete if any word in  $M$  is a factor of some word in  $X^*$  [J. Néraud, C. Selmi, Free monoid theory: maximality and completeness in arbitrary submonoids, *Internat. J. Algorithms Comput.* 13(5) (2003) 507–516]. Given a code  $X \subseteq M$ , we are interested in the construction of a weakly  $M$ -complete code that contains  $X$ , if it exists. In the case where  $M$  and  $X$  are regular sets, the existence of such a code has been established [J. Néraud, Completing a code in a regular submonoid of the free monoid, in acts of MCU'2004, *Lecture Notes in Computer Sciences*, Vol. 3354, Springer, Berlin, 2005, pp. 281–291; J. Néraud, On the completion of codes in submonoids with finite rank, *Fund. Inform.*, to appear]. Actually, this result lays upon a method of construction that preserves the regularity of sets. As well known, any regular (or finite) code may be embedded into a regular (finite) prefix code that is complete in  $A^*$ . In the framework of the weak completeness, we prove that the following problem is decidable:

*Instance:* A regular submonoid  $M$  of  $A^*$ , and a regular (or finite) prefix code  $X \subseteq M$ .

*Question:* Does a weakly  $M$ -complete regular (finite) prefix code containing  $X$  exist?

© 2006 Elsevier B.V. All rights reserved.

MSC: 05E99; 68R15; 68Q45; 94A45

*Keywords:* Combinatorial problem; Formal language; Words; Free monoid; Submonoid; Maximal code; Dense; Complete; Weak completeness; Completion of codes; Regular prefix codes

## 1. Introduction

Completeness of codes is one of the most important concepts of the free monoid theory: this is due to its mathematical relevance just as much as its potential applications, particularly in the framework of the theory of information. Given an alphabet  $A$ , and given a subset  $X$  of  $A^*$ ,  $X$  is complete if and only if any word of  $A^*$  is a factor of some word in  $X^*$ , the submonoid of  $A^*$  which is generated by  $X$ . Taking into account of the numerous challenging problems that are connected to this notion, it is not surprising that a lot of different concepts of local completeness has been introduced (cf. e.g. [3,4,9]).

Our paper deals with the so-called concept of “weak completeness” in a submonoid of  $A^*$ : given an arbitrary submonoid  $M$  of  $A^*$ , and given a subset  $X$  of  $M$ ,  $X$  is weakly  $M$ -complete if and only if any word in  $M$  is a factor of some word of  $X^*$ . The introduction of this notion is particularly justified by the fact that, as explained in [8,9], given the language  $L(S)$  of the finite blocks of a sofic system  $S$ , studying the local notion of completeness with respect to

E-mail address: [Jean.Neraud@univ-rouen.fr](mailto:Jean.Neraud@univ-rouen.fr)

URL: <http://neraud.jean.free.fr>.

$L(S)$  comes down to examine weak completeness in particular monoids that are included in  $L(S)$ , i.e. the so-called stabilizers of the states in the automaton with behavior  $S$ .

According to Zorn's lemma, given an arbitrary submonoid  $M$  of  $A^*$ , and given a code  $X \subseteq M$ , the family of the codes  $\hat{X}$  that satisfy  $X \subseteq \hat{X} \subseteq M$  has a maximal element. This property, conjugated with a result in [6], allows to establish the existence, for any code  $X \subseteq M$ , of a weakly  $M$ -complete code  $\hat{X}$  such that  $X \subseteq \hat{X} \subseteq M$ . However, in the most general case, exhibiting a method of construction that is, for short, “completing”  $X$  in  $M$ , remains an open question.

In [5], we present a general formula which, in particular, allows to complete a code in a regular submonoid. Despite the fact that this method preserves the regularity of codes, it does not bring new informations concerning the problem of completion in the framework of special classes of codes, such as prefix codes, or codes with a bounded deciphering delay. However, elaborating such a method of completion for a regular (or finite) prefix code in a regular submonoid is of peculiar interest: indeed, most of the coding algorithms (cf. e.g. [11]) lay upon efficient methods which allow effectively to embed a prefix code  $X \subseteq A^*$  in a complete one.

In our paper, starting from an arbitrary submonoid  $M$  of  $A^*$ , we consider a prefix code  $X \subseteq M$ . We give a characterization of the existence of a weakly  $M$ -complete code  $\hat{X}$  that contains  $X$ . In the case where  $X$  is a regular (or a finite) prefix code, we show that the existence of  $\hat{X}$  is a decidable problem, moreover, we present a formula for computing such a code  $\hat{X}$  when it exists. Actually, our method lays upon the investigation of the structure of  $M$  itself. In particular, we make use of the examination of the properties of a special “test prefix code” in  $M$ .

We now present the contents of the paper:

Section 2 contains the preliminaries: the terminology of the free monoid is preferably settled. Given a submonoid  $M$  we introduce the so-called notion of weak  $M$ -completeness. We recall some of our results concerning this notion, especially in the case of prefix codes: they will be useful in the sequel of the paper.

In Section 3, we produce a theoretical study. In the most general case, two groups of formula are presented for obtaining a characterization of the existence of the preceding set  $\hat{X}$ , and for constructing  $\hat{X}$  itself. Actually, the first group of formula introduces canonical sets in the submonoid  $M$ , namely,  $M_0$  and  $Y_0$ . This leads to characterize the existence of a so-called strongly  $M$ -thin code  $\hat{X}$ .

The existence and the computation itself of a regular, or finite, prefix code  $\hat{X}$  are examined in the Sections 4 and 5.

Different remarks and perspectives are drawn in the conclusion.

## 2. Preliminaries

We adopt the standard notations of the free monoid theory: given a word  $w$  in  $A^*$  (the free monoid generated by  $A$ ), we denote by  $|w|$  its length, the empty word being the word of length zero.

Given two words  $u, w \in A^*$ , we say that  $u$  is a *factor* (*prefix*, *suffix*) of  $w$  if and only if we have  $w \in A^*uA^*$  ( $w \in uA^*$ ,  $w \in A^*u$ ). Given a subset  $X$  of  $A^*$ , we denote by  $F(X)$  ( $P(X)$ ,  $S(X)$ ), the set of the words that are factor (prefix, suffix) of some word in  $X$ .

### 2.1. Completeness in an arbitrary submonoid of $A^*$

We assume the reader to be familiar with the basic concepts from the theory of variable length codes, and we suggest, if necessary, that he (she) refers to [1], which constitutes a classical reference in the domain. We simply recall that given a code  $X \subseteq A^*$ ,  $X$  is *prefix* if and only if  $X \cap XA^+ = \emptyset$  or, equivalently, if  $X^*$  is a (right)-unitary submonoid of  $A^*$ , that is  $(X^*)^{-1}X^* = X^*$ . In the topic of coding algorithms complete prefix codes play a prominent part [11]. More precisely, codes  $X \subseteq A^*$  are required in such a way that any word in  $A^*$  is a prefix of some word in  $X^*$ .

Let  $M$  be an arbitrary submonoid of  $A^*$ , and let  $X \subseteq M$ . We say that  $X$  is *weakly  $M$ -complete* if and only if any word in  $w \in M$  is a factor of some word in  $X^*$ , i.e.  $A^*wA^* \cap X^* \neq \emptyset$ . Note that this notion of completeness is more general than the classical notion of  $M$ -completeness in monoids, where the condition  $MwM \cap X^* \neq \emptyset$  must be satisfied by any word  $w \in M$ .

**Example 1.** Let  $A = \{a, b\}$ ,  $M = bA^*$ , and  $X = \{b^{|w|}aw : w \in A^+\}$ .  $X$  is a prefix code. Indeed, assuming that  $b^{|w|}aw \in P(b^{|w'|}aw')$ , we have  $|w| = |w'|$  and  $w \in P(w')$ , thus  $w = w'$ .

Let  $v \in M$ , and let  $i$  be the greatest positive integer such that  $v \in b^i A^*$ . If  $i = |v|$  then, trivially,  $v$  is a prefix of  $b^{|v|}ab^{|v|} \in X^*$ . Assuming that  $v = b^i aw$ , if  $|w| \leq i$ , then we have  $v \in P(b^i awb^{i-|w|}) \subseteq P(X)$ , otherwise the word  $v$  is a suffix of  $b^{|w|}aw$ . Consequently we obtain  $M \subseteq F(X)$ , hence  $X$  is weakly  $M$ -complete.

Actually, from the point of view of coding the elements of  $M$ , the factorization of words over the prefix code  $X$  has a poor inference. Indeed, words in  $M \setminus P(X^*)$  exist, moreover, we have  $M \subseteq P(X) \cup S(X)$ .

As a matter of fact, in [2, Proposition 2.5], the existence of a weakly  $M$ -complete code is stated:

**Theorem 2.** *Given an arbitrary submonoid  $M$  of  $A^*$ , a weakly  $M$ -complete prefix code exists.*

The proof of Theorem 2 leads effectively to construct a weakly  $M$ -complete code, say  $Y$ . However, in any case, a word  $w \in M \setminus P(Y^*)$  exists, an hitch for potential applications of these result in the topic of coding algorithm.

From this point of view, we are interested by the construction of prefix codes  $X \subseteq M$ , such that any word in  $M$  is a prefix of some word in  $X^*$ .

In the next, we shall present a large class of prefix codes which satisfy this property.

**Definition 3.** Given a submonoid  $M \subseteq A^*$ , and given a set  $X \subseteq M$ ,  $X$  is strongly  $M$ -thin (s. $M$ -thin for short) if we have  $M \setminus F(X) \neq \emptyset$ .

In [6], the specificity of strong thinness is emphasized by comparisons with other classes of sets that are connected to thinness, such as thin sets or very thin sets [1].

Given an arbitrary submonoid  $M \subseteq A^*$ , regular codes  $X \subseteq M$  constitute a large class of s. $M$ -thin codes. Indeed, according to [1, p. 224], a word  $w \in X^* \setminus F(X)$  exists, thus we have  $M \setminus F(X) \neq \emptyset$ .

One of the most remarkable properties of s. $M$ -thin codes is stated by the following result [6, Theorem 4]:

**Theorem 4.** *Let  $X \subseteq M$  be a s. $M$ -thin prefix code. Then the three following conditions are equivalent:*

- (i)  *$X$  is weakly  $M$ -complete (i.e.  $A^*wA^* \cap X^* \neq \emptyset$ , for any word  $w \in M$ ).*
- (ii)  *$X$  is  $M$ -complete (i.e.  $MwM \cap X^* \neq \emptyset$ , for any word  $w \in M$ ).*
- (iii)  *$M$  is a subset of  $P(X^*)$ .*

Despite the result of Theorem 2, given an arbitrary submonoid  $M \subseteq A^*$ , the existence of a s. $M$ -thin  $M$ -complete prefix code remains an open problem. The aim of our paper consists in studying the following question:

**Problem 5.** *Instance:* A submonoid  $M$  of the free monoid  $A^*$ , and  $X \subseteq M$  a s. $M$ -thin prefix code.

*Question:* Does a s. $M$ -thin  $M$ -complete prefix code that contains  $X$  exist?

In fact, the preceding question may be declinated in asking for the existence of a regular, even finite,  $M$ -complete prefix code. Clearly, given an instance of Problem 5, an affirmative answer allows to state the existence itself of a  $M$ -complete prefix code.

In such a context, any prefix set  $\hat{X}$  will be called a *solution* of Problem 5 with instance  $(X, M)$ . In Section 4, in the case where  $X$  and  $M$  are regular sets, we shall prove that Problem 5 is decidable, by presenting the scheme of an algorithm that computes such a solution. Before this, in Section 3, we produce a theoretical study.

### 3. A test set for the completion of prefix codes in a submonoid

We begin by stating some necessary and sufficient conditions for the existence of a solution for Problem 5. Note that in the two next subsections, the results that we obtain are stated without the assumption that the prefix codes  $X$ , and  $\hat{X}$  are s. $M$ -thin.

#### 3.1. Existence of solutions: some necessary conditions

The first property that we state expresses, in a straightforward way, the fact that  $\hat{X}^*$  is a right unitary submonoid of  $A^*$ :

**Lemma 6.** Let  $M$  be an arbitrary submonoid of  $A^*$ , and let  $X \subseteq M$  be a prefix code. Assume that a prefix code  $\hat{X}$  exists such that the following condition holds:

$$X \subseteq \hat{X} \subseteq M \subseteq P(\hat{X}^*).$$

Then we have  $X^{-1}M \subseteq P(M)$ .

**Proof.** We have  $X^{-1}M \subseteq \hat{X}^{-1}P(\hat{X}^*) \subseteq P(\hat{X}^*) \subseteq P(M)$ .  $\square$

**Example 7.** Let  $A = \{a, b\}$ , and  $M = (a^*ba^*)^*$ . Take  $X = \{b, ab\}$  and  $\hat{X} = a^*b$ . The prefix codes  $\hat{X}$  and  $X$  satisfies the conditions of Lemma 6. Indeed, let  $w = a^i ba^j w' \in M$ , with  $w' \in M$ , thus  $a^j w' \in M$ .

Since we have  $X^{-1}w = \{a^j w'\} \subseteq M$ , in any case we obtain  $X^{-1}M \subseteq M$ . In another side, it follows by induction on  $|w|$  that we have  $w \in P(\hat{X}^*)$ .

Now, we present the construction of a special prefix code, namely  $Y_0$ , whose properties will allow to solve Problem 5:

$$\begin{aligned} M_0 &= \{t \in M : t^{-1}P(M) \subseteq P(M)\}, \\ Y_0 &= (M_0 \setminus M_0A^+) \setminus \{\varepsilon\}. \end{aligned} \tag{1}$$

By construction  $Y_0$  is a prefix set, moreover, it satisfies the following result:

**Lemma 8.** With the preceding notation, the three following properties hold:

- (i)  $M_0$  is submonoid of  $M$ .
- (ii) For any prefix code  $Y \subseteq M$  if we have  $M \subseteq P(Y^*)$ , then we have  $Y^* \subseteq M_0$ .
- (iii) We have  $P(M_0) \subseteq P(Y_0^*)$ .

**Proof.** (i) Clearly, we have  $\varepsilon \in M_0$ . For all  $t, t' \in M_0$ , we have:

$$(tt')^{-1}P(M) = t'^{-1}(t^{-1}P(M)) \subseteq t'^{-1}P(M) \subseteq P(M),$$

therefore,  $M_0$  is closed under concatenation and a submonoid of  $M$ .

(ii) Since  $Y^*$  is a unitary submonoid of  $A^*$ , we have:

$$(Y^*)^{-1}P(M) \subseteq (Y^*)^{-1}P(Y^*) \subseteq P(Y^*) \subseteq P(M),$$

and thus, by definition,  $Y^* \subseteq M_0$ .

(iii) Let  $w \in P(M_0)$ . By construction, we have  $w = x_0 w'$ , with  $x_0 \in Y_0 \subseteq M_0$ , and  $w' \in M_0^{-1}M_0$ , thus,  $w' \in P(M_0)$ . With such a condition, Property (iii) from our statement follows by induction on the length of  $w$ .  $\square$

Given a language  $Z \subseteq A^*$ , its stabilizer is  $Stab(Z) = \{u \in A^* : uZ \subseteq Z\}$ . Property (ii) from Lemma 8 expresses a fact that is similar to the following result from [3, Lemma 3.9]:

**Theorem 9.** Given a submonoid  $M \subseteq A^*$ ,  $Stab(P(M))$  is the greatest submonoid  $U^*$  of  $A^*$  which satisfies  $M \subseteq P(U^*)$ .

In Section 5, we shall see that the two preceding concepts intersect in the case of finite prefix codes.

### 3.2. Toward the characterization of solutions for Problem 5

Let  $(X, M)$  be an instance of Problem 5. Given a prefix code  $Z \subseteq M$ , we introduce the following sets:

$$\begin{aligned} S_Z &= Z^* \setminus P(XA^*), \\ Y_Z &= S_Z \setminus S_Z \cdot Z^+, \\ \hat{X}_Z &= X \cup Y_Z. \end{aligned} \tag{2}$$

According to Lemma 8, if we have  $M \subseteq P(Z^*)$ , then we obtain  $Z^* \subseteq M_0$ , moreover, the following property holds:

**Lemma 10.** *With the preceding notation, we have  $S_Z = Y_Z \cdot Z^*$ .*

**Proof.** Let  $w = yz$ , with  $y \in Y_Z$  and  $z \in Z^*$ , thus  $w \in Z^*$ . Assuming that we have  $w \notin S_Z$ , we obtain  $w \in P(XA^*)$ , thus  $y \in P(XA^*)$ , a contradiction with  $y \in Y_Z$ . Consequently, we have  $Y_Z \cdot Z^* \subseteq S_Z$ . Since we have either  $y \in P(x)$  or  $x \in P(y)$ , we obtain  $y \in P(XA^*)$ , a contradiction with  $y \in Y_Z$ . Consequently, we have  $Y_Z \cdot Z^* \subseteq S_Z$ .

Let  $w \in S_Z \subseteq Z^*$ . Since  $Z$  is a prefix code, a unique sequence of words in  $Z$ , namely,  $z_1, \dots, z_n$ , exists such that  $w = z_1 \dots z_n$ . Denote by  $k$  the smallest integer in  $[1, n]$  such that  $z_1 \dots z_k \in S_Z$  (such an integer exists, since we have  $w = z_1 \dots z_n \in S_Z$ , thus  $w \notin P(XA^*)$ ). Let  $z$  be a prefix of  $z_1 \dots z_k$  which belongs to  $S_Z$ . Since  $Z$  is a prefix code, and from the minimality of  $k$ , we have  $z_1 \dots z_k \in Y_Z$  thus  $w \in Y_Z \cdot Z^*$ .  $\square$

As a matter of fact, the following result brings a characterization for the existence of a prefix code  $\hat{X}$  such that the inclusion  $X \subset \hat{X} \subseteq M \subseteq P(\hat{X}^*)$  holds.

**Lemma 11.** *Let  $M \subseteq A^*$  be an arbitrary submonoid of  $A^*$ , and let  $X \subseteq M$  be a prefix code. The three following properties are equivalent:*

- (i)  *$X$  is embeddable in a prefix code  $\hat{X}$  such that  $X \subseteq \hat{X}$  and  $M \subseteq P(\hat{X}^*)$ .*
- (ii) *A prefix code  $Z \subseteq M$  exists such that  $M \subseteq P(Z^*)$  and  $X^{-1}M \subseteq P(M)$ .*
- (iii) *We have  $M \subseteq P(Y_0^*)$  and  $X^{-1}M \subseteq P(M)$ .*

**Proof.** Let  $X \subseteq M$  be a prefix code. Assume that Condition (i) holds. It follows from Lemma 8 that we have  $\hat{X}^* \subseteq M_0$  and  $M_0 \subseteq P(M_0) \subseteq P(Y_0^*)$ . This implies  $M \subseteq P(\hat{X}^*) \subseteq P(Y_0^*)$ . Moreover, according to Lemma 6, we have  $X^{-1}M \subseteq P(M)$  and Condition (iii) follows.

Trivially, Condition (iii) implies Condition (ii). It remains to prove that Condition (ii) implies Condition (i). Assuming that a prefix code  $Z$  exists such that  $M \subseteq P(Z^*)$  and  $X^{-1}M \subseteq P(M)$ , we consider the set  $\hat{X}_Z$  as constructed in (2).

- (1) First, we prove that  $\hat{X}_Z$  is a prefix code. Let  $u, v$  be two words in  $\hat{X}_Z$ . Assume that  $v$  is a proper prefix of  $u$ . According to the definition of  $\hat{X}_Z$ , exactly one of the four following cases occurs:
  - *Condition  $u, v \in X$ .* Since  $X$  is a prefix set, this condition cannot hold.
  - *Condition  $u \in X, v \in Y_Z$ .* We have  $v \in Y_Z \subseteq Z^*$  and  $v \in P(u) \subseteq P(X) \subseteq P(XA^*)$  thus, by definition,  $v \notin S_Z$ : a contradiction with  $Y_Z \subseteq S_Z$ .
  - *Condition  $u \in Y_Z, v \in X$ .* With this condition, we have  $u \in vA^* \subseteq XA^* \subseteq P(XA^*)$  thus, by definition  $u \notin S_Z$ . As in the preceding case, this contradicts the definition of  $Y_Z$ .
  - *Condition  $u, v \in Y_Z$ .* By definition, we have  $u, v \in S_Z \subseteq Z^*$ . Since  $Z^*$  is a right unitary submonoid, this implies  $v^{-1}u \in Z^*$ . It follows from  $v \neq u$  that  $u \in vZ^+ \subseteq S_Z Z^+$ , thus  $Y_Z \cap S_Z Z^+ \neq \emptyset$  which, once more, contradicts the definition of  $Y_Z$ .

In each case, we obtain a contradiction, hence  $\hat{X}_Z$  is a prefix code.

- (2) Now, we prove that  $M \subseteq P(\hat{X}_Z^*)$ . Let  $w \in P(M)$ . According to Condition (ii) from the statement, we have  $w \in P(Z^*)$ . Trivially, if  $w$  is a prefix of some word in  $X^*$ , then we have  $w \in P(\hat{X}_Z^*)$ . Assuming that we have  $w \notin P(X^*)$  we denote by  $x$  the longest word in  $X^*$  such that  $x \in P(w)$ , and we set  $w' = x^{-1}w$ .

If we have  $x = \varepsilon$  then, by definition, we obtain  $w \in P(S_Z)$ . Otherwise, according to Condition (ii), we have  $X^{-1}M \subseteq P(M)$ , hence we obtain  $w' = x^{-1}w \in P(M) \subseteq P(Z^*)$ , thus, by construction  $w' \in P(S_Z)$ . Now, it follows from Lemma 10 that in any case, we have  $w' \in P(Y_Z \cdot Z^*)$ .

If  $w'$  belongs to  $P(Y_Z)$  then we obtain  $w \in X^*P(Y_Z) \subseteq P(\hat{X}_Z^*)$ . Otherwise, we have  $w' = yw_1$ , with  $y \in Y_Z$ , and  $w_1 \in P(Z^*) \subseteq P(M)$ . Consequently we obtain  $w \in X^* \cdot Y_Z \cdot P(M) \subseteq \hat{X}_Z^* \cdot P(M)$ , and the inclusion  $M \subseteq P(\hat{X}_Z^*)$  will follow by induction on the length of  $w \in P(M)$ .

This completes the proof of Lemma 11.  $\square$

### 3.3. Answering to Problem 5

In the case where  $\hat{X}$  is required to be a s.M-thin set, the following result interprets the contents of Lemma 11:

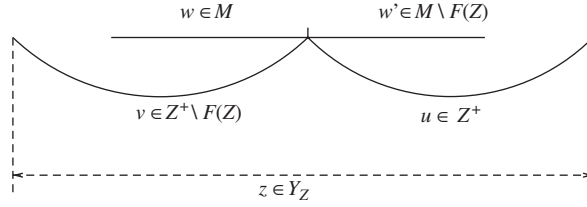


Fig. 1. Proof of Theorem 12.

**Theorem 12.** Let  $M$  be a submonoid of  $A^*$  and let  $X \subseteq M$  be  $s.M$ -thin prefix code. With the preceding notation, the three following conditions are equivalent:

- (i)  $X$  is embeddable in a  $s.M$ -thin  $M$ -complete prefix code.
- (ii) A  $M$ -complete  $s.M$ -thin prefix code  $Z$  exists and we have  $X^{-1}M \subseteq P(M)$ .
- (iii)  $Y_0$  is a  $M$ -complete  $s.M$ -thin prefix code and we have  $X^{-1}M \subseteq P(M)$ .

**Proof.** (a) Assume that Condition (i) holds, and let  $\hat{X}$  be a  $M$ -complete  $s.M$ -thin prefix code containing  $X$ . According to Theorem 4 we have  $X \subseteq \hat{X}$  and  $M \subseteq P(\hat{X}^*)$ , thus the Condition (i) of Lemma 11 holds. As a consequence, we have  $M \subseteq P(Y_0^*)$  and  $X^{-1}M \subseteq P(M)$ .

We shall prove that  $Y_0$  is a  $s.M$ -thin set. Let  $w \in M$  such that a word  $v \in Y_0$  exists with  $w \in F(v)$ . Since  $\hat{X}$  is  $M$ -complete, we have  $v \in M \subseteq P(\hat{X}^*)$ . According to the property (ii) of Lemma 8, we have  $\hat{X} \subseteq M_0$ . Assuming  $v \in \hat{X}A^+$ , we obtain  $v \in Y_0 \cap M_0A^+$ , thus  $v \in M_0 \cap M_0A^+$ , a contradiction with  $v \in Y_0$ . This implies  $v \in P(\hat{X})$ , thus  $w \in F(\hat{X})$ .

Consequently, since  $\hat{X}$  is a  $s.M$ -thin set, a word  $w \in M \setminus F(Y_0)$  exists, thus,  $Y_0$  is a  $s.M$ -thin set: we obtain Condition (iii) of our theorem. Trivially, Condition (ii) follows.

(b) Now, we assume that Condition (ii) holds, and we consider the set  $\hat{X}_Z$  as defined in Construction (2). According to Lemma 11,  $\hat{X}_Z$  is a prefix code such that  $X \subset \hat{X}_Z$  and we have  $M \subseteq P(\hat{X}_Z^*)$ . It remains to prove that  $\hat{X}$  is  $s.M$ -thin.

Since  $Z$  is  $s.M$ -thin, a word  $w' \in M \setminus F(Z)$  exists. Since  $X$  is a  $s.M$ -thin set, a word  $w \in M \setminus P(X)$  exists. By contradiction, we assume that a word  $z \in Y_Z$  exists such that we have  $ww' \in F(z)$ .

According to Theorem 4, we have  $z \in Y_Z \subseteq P(Z^*)$ . Since we assume  $w' \in M \setminus F(Z)$ , a pair of words  $v, u \in Z^+$  exists such that we have  $z = vu$ , with  $w \in F(v)$  (cf. Fig. 1).

Assuming that we have  $v \in S_Z$ , we obtain  $z \in S_Z Z^+$ , which contradicts  $z \in Y_Z$ . As a consequence, we have  $v \in Z^* \setminus S_Z$ . Since we have  $w \in M \setminus P(X)$ , this implies  $v \in P(XA^*)$ , which contradicts  $z \in Y_Z$ . As a consequence, no word  $z \in Y_Z$  may exist such that we have  $ww' \in F(z)$ , that is  $ww' \notin F(Y_Z)$ .

In a similar way, since  $X$  is  $s.M$ -thin, a word  $w'' \in M \setminus F(X)$  exists. Since we have  $ww'w'' \in M \setminus F(X \cup Y_Z)$ , the prefix code  $\hat{X}_Z = X \cup Y_Z$  is a  $s.M$ -thin set. This completes the proof of Theorem 12.  $\square$

#### 4. The case of regular prefix codes

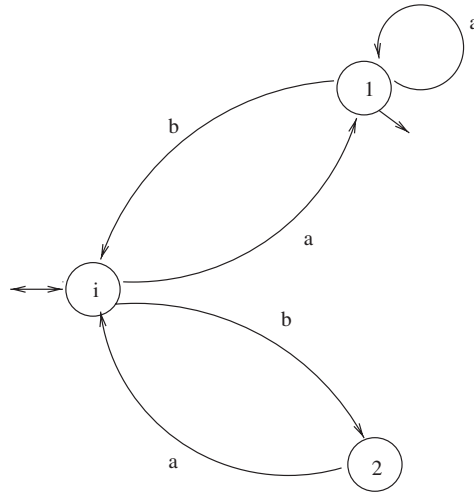
In the case where  $M$  and  $X$  are assigned to be regular, Theorem 12, may be declinated as indicated in the following:

**Theorem 13.** Let  $M$  be a regular submonoid of  $A^*$  and let  $X \subseteq M$  be a regular prefix code. With the preceding notation, the three following conditions are equivalent:

- (i)  $X$  is embeddable in a regular  $M$ -complete prefix code.
- (ii) A regular  $M$ -complete prefix code exists and we have  $X^{-1}M \subseteq P(M)$ .
- (iii)  $Y_0$  is a regular  $M$ -complete prefix code and we have  $X^{-1}M \subseteq P(M)$ .

**Proof.** (a) First, we prove that Condition (i) implies Condition (iii). Since any regular code is  $s.M$ -thin (cf. Section 2), according to Theorem 12,  $Y_0$  is a  $M$ -complete prefix code, moreover, we have  $X^{-1}M \subseteq P(M)$ . For establishing that



Fig. 2. Example 15:  $T_0 = \{i\}$ .

$Y_0$  is a regular set we shall prove that  $M_0$  is regular.

Actually, we prove that the following property holds:

$$u^{-1}P(M) = v^{-1}P(M) \implies u^{-1}M_0 = v^{-1}M_0. \quad (3)$$

Since  $M$  itself is regular, this will imply that the syntactic monoid of  $M_0$  is finite.

Assuming that we have  $u^{-1}P(M) = v^{-1}P(M)$ , we consider a word  $t \in u^{-1}M_0$ , and we prove that  $t \in v^{-1}M_0$ . For any word  $s \in A^*$  such that  $v(ts) \in P(M)$ , we have also  $u(ts) \in P(M)$ . Since we have  $X^{-1}M \subseteq P(M)$ , from  $ut \in M_0$ , it follows that  $s \in P(M)$ , thus  $vt \in M_0$  and  $t \in v^{-1}M_0$ .

Consequently, Condition (i) implies Condition (iii). Trivially Condition (ii) follows.

(b) Assume that Condition (ii) holds, and let  $Z$  be a regular  $M$ -complete prefix code. According to Theorem 12, and since the construction in (2) preserves the regularity of sets, Condition (i) follows.  $\square$

We assume the reader to be familiar with the basic notions from automata theory and, for any complement of information, we suggest that he (she) refers to [10]. We consider the minimal trim automaton with behavior  $M$ . Denote by  $Q$  the set of states, by  $i$  the initial state, and by  $T$  the set of the terminal states. Given a path  $(q, w, q')$  in the automaton, we set  $q' = q \cdot w$ . More generally, given a subset  $E$  of  $A^*$ , and given a state  $q$ , we set  $q.E = \{q' : \exists w \in E \ q' = q \cdot w\}$ . According to the preceding property (3), for any pair of words  $w, w_0 \in M$ , the equality  $w^{-1}M = w_0^{-1}M$  implies  $w^{-1}M_0 = w_0^{-1}M_0$ . As a consequence, a set  $T_0 \subseteq T$  exists such that  $T_0 = i.M_0$ . This leads to a decision algorithm for solving Problem 5 in the case of an instance composed of regular sets. In the following, we indicate the scheme of a corresponding procedure, which compute a solution of Problem 5 when it exists.

#### Algorithm (A).

(Step 1)  $T_0 \leftarrow \{t \in T : \forall w \in A^*, t.w \in T \implies i.w \in Q\}$ .

(Step 2)  $M_0 \leftarrow \{w \in A^* : i.w \in T_0\}$

(Step 3)  $Y_0 \leftarrow (M_0 \setminus M_0A^+) \setminus \{\varepsilon\}$

(Step 4) **if**  $M \subseteq P(Y_0^*)$  **and**  $i \cdot X \subseteq T_0$  **then**

$Z \leftarrow Y_0$ ; compute  $\hat{X}$  by applying Formula (2)

**else return** ("no completion for  $X$ ")

**endif**

**endalgorithm**

The condition  $i \cdot X \subseteq T_0$  is equivalent to  $X^{-1}M \subseteq P(M)$ . Moreover, each of the preceding steps of computation will be done by applying elementary operations on finite automata, such as extraction or comparison of subgraphs (cf. e.g. [10, p. 69]).

**Corollary 14.** *Let  $M$  be a regular submonoid of  $A^*$  and let  $X \subseteq M$  be a regular prefix code. The existence of a regular  $M$ -complete prefix code containing  $X$  is decidable.*

**Example 15.** Let  $M = \{a, ab, ba\}^*$ . From the minimal automaton with behavior  $M$ , we obtain  $T_0 = \{i\}$ .

$M_0$  is the stabilizer of  $\{i\}$  and we have  $Y_0 = \{a^k b : k \geq 1\} \cup \{ba\}$ . Moreover, we have  $M \subseteq P(Y_0^*)$  (cf. Fig. 2).

The prefix code  $X = \{baba\}$ , satisfies  $X^{-1}M \subseteq P(M)$ . It may be embedded into the prefix code  $\hat{X} = \{baba\} \cup [(\{\varepsilon, ba\}\{a^k b : k \geq 1\}^*) \setminus \{\varepsilon\}]$ .

## 5. The case of finite prefix codes

In [4, Proposition 3.5], a characterization of the existence of a finite  $M$ -complete code is given:

**Theorem 16.** *Let  $M$  be an arbitrary submonoid of  $A^*$ . Then the three following conditions are equivalent:*

- (i) *A finite  $M$ -complete prefix code exists.*
- (ii)  *$M^{-1}M \subseteq P(M)$ .*
- (iii)  *$(M \setminus MA^+) \setminus \{\varepsilon\}$  is a finite  $M$ -complete prefix set.*

According to Condition (ii) of Theorem 16, we have in fact  $M_0 = M$ , thus,  $Y_0 = (M \setminus MA^+) \setminus \{\varepsilon\}$ . As a consequence, the following statement may be formulated:

**Theorem 17.** *Let  $M$  be an arbitrary submonoid of  $A^*$  and let  $X \subseteq M$  be a finite prefix code. Then the four following conditions are equivalent:*

- (i) *A  $M$ -complete finite prefix code that contains  $X$  exists.*
- (ii)  *$Y_0$  is a  $M$ -complete finite prefix code.*
- (iii) *We have  $M^{-1}M \subseteq P(M)$ .*
- (iv) *A  $M$ -complete finite prefix code exists.*

**Proof.** (a) According to Theorem 16, Conditions (ii)–(iv) are equivalent. Assuming that Condition (i) holds, according to Theorem 13, Condition (ii) follows.

(b) Now, we assume that Condition (iv) holds, that is, a  $M$ -complete finite prefix code  $Z$  exists. According to Theorem 16,  $X$  satisfies the condition  $X^{-1}M \subseteq P(M)$ . It remains to prove that  $\hat{X}_Z$  is a finite set, with respect to the notation in (2).

Given a word  $y \in Y_Z \subseteq Z^*$ , denote by  $t$  be the longest prefix of  $y$  which belongs to  $P(X)$ , and let  $\alpha$  be the shortest word such that  $t\alpha \in Z^*$ . According to the definition of  $S_Z$ , no prefix of the word  $y$  may belong to  $X$ , therefore, we have  $t\alpha \in Y_Z$ . This implies  $y = t\alpha$ .

As a consequence, we have  $\hat{X}_Z \subseteq X \cup P(X) \cdot S(Z)$ , therefore,  $\hat{X}$  is a finite set.  $\square$

Taking account of Theorem 17, in the case where the submonoid  $M$  is generated by a finite set, Algorithm (A) may be simplified as indicated in the following:

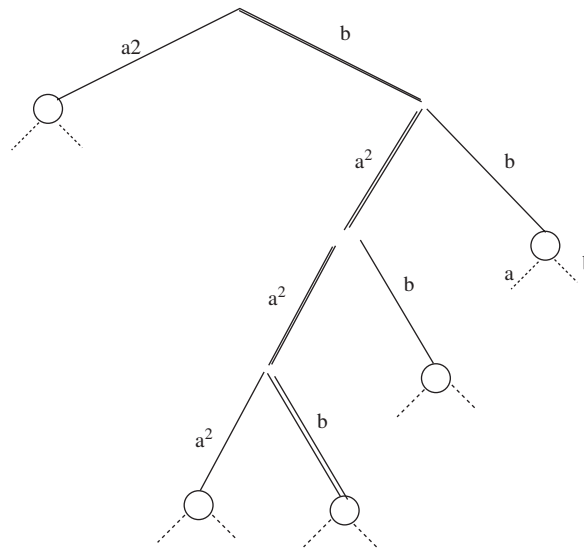
**Algorithm (B).**

```
(Step 1)  $Y_0 \leftarrow (M \setminus MA^+) \setminus \{\varepsilon\}$ 
(Step 2) if Condition (ii) from Theorem 17 holds then
     $Z \leftarrow Y_0$ 
    compute  $\hat{X}_Z$  by applying Formula (2)
else (“no completion for  $X$ ”)
endif
endalgorithm
```

**Remark 18.** (1) According to [7, Theorem 12], a  $O((\sum_{y \in Y_0} |y|)^2)$  procedure may be implemented for deciding whether Condition (ii) from Theorem 17 holds.

(2) The result of Theorem 17 expresses a characterization which makes only use of the structure of the submonoid  $M$ .





(3) It is also of interest to note that, if Condition (iii) in Theorem 17 does not holds, according to Theorem 13, no finite prefix code  $X \subseteq M$  may be embedded in a regular  $M$ -complete prefix code.

We have  $Y_0 = \{a^2, b\}$ . Let  $X = \{ba^4b\}$ .

With the preceding notation, we have:

$$S_{Y_0} = \{a^2, b^2, ba^2b, ba^6\}Y_0^*.$$

This implies:

$$\hat{X} = X \cup Y_{Y_0} = \{ba^4b, a^2, b^2, ba^2b, ba^6\}.$$

$Y_{Y_0}$  is a finite prefix  $M$ -complete code. For instance, the word  $b^3 a^{16} b^3 \in M$  is a prefix of  $(b^2)(ba^6)(a^2)^5(b^2)(ba^2b) \in Y^*$ .

## 6. Concluding remarks

The results that we have obtained in this paper lead to a break with those of the classical theory of codes, where it is well known that any regular (or finite) prefix code is embeddable in a complete one. In particular, regular submonoids of  $A^*$  exist such that no regular prefix code has a regular completion. In another side, according to Néraud [5], in such a submonoid, any regular code has a regular completion. A natural question consists in drawing similar studies in the framework of special classes of codes, such as circular codes, or codes with a bounded deciphering delay. Clearly, the question of the existence of finite complete codes in an arbitrary submonoid of  $A^*$  is also concerned: it turns up to the question of the finite completion, one of the most famous and difficult problem from the theory of codes.

## References

- [1] J. Berstel, D. Perrin, *Theory of Codes*, Academic Press, New York, 1985.
- [2] F. Blanchard, G. Hansel, systèmes codés, *Theoret. Comput. Sci.* 44 (1986) 17–49.
- [3] M. Latteux, E. Timmerman, Finitely generated  $\omega$ -languages, *Inform. Process. Lett.* 23 (1986) 171–175.

- [4] I. Litovsky, Prefix-free languages as  $\omega$ -generators, *Inform. Process. Lett.* 37 (1991) 61–65.
- [5] J. Néraud, Completing a code in a regular submonoid of the free monoid, in acts of MCU'2004, *Lecture Notes in Computer Science*, Vol. 3354, Springer, Berlin, 2005, pp. 281–291; J. Néraud, On the completion of codes in submonoids with finite rank, *Fund. Inform.*, to appear.
- [6] J. Néraud, C. Selmi, Free monoid theory: maximality and completeness in arbitrary submonoids, *Internat. J. Algorithms Comput.* 13 (5) (2003) 507–516.
- [7] J. Néraud, C. Selmi, A characterization of complete finite prefix codes in arbitrary submonoids of  $A^*$ , *J. Aut. Langu. Comb.* 9 (1) (2004) 103–110.
- [8] A. Restivo, Finitely generated sofic systems, *Theoret. Comput. Sci.* 65 (1989) 55–64.
- [9] A. Restivo, Codes and local constraints, *Theoret. Inform. Appl.* 72 (1990) 265–270.
- [10] J. Sakarovitch, *Éléments de la théorie des automates*, Vuibert Inform. (2003).
- [11] J. Ziv, A. Lempel, A universal algorithm for sequential data compression, *IEEE Trans. Inform. Theory* IT-23 (1977) 337–343.